社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.
HIP97-2 (1997-06)

顔表情のガボール・ウエイブレット表現

マイケル　ライオンズ[1]、　蒲池　みゆき[2]、　行場　次朗[2]、　赤松　茂[1]

[1] ＡＴＲ人間情報通信研究所　京都府相楽郡精華町光台２－２
[2] 九州大学文学部　福岡市東区箱崎６－１９－１

{mlyons, akamatsu} @hip.atr.co.jp, {kamaltm, gyobaltm} @mbox.nc.kyushu-u.ac.jp

あらまし　ガボール・ウエイブレットを取り入れた顔パターンの符号化方法に関して、顔表情の知覚モデルとしての妥当性を検討した。ガボール・ウエイブレットフィルターは、大脳皮質視覚領野でなされる顔画像符号化のモデルとして導入されたものである。一方、人間の表情知覚は、意味微分法（ＳＤ法）を用いて測定した。モデルにより算出された表情類似度と、実際に測定した類似度は有意な高い相関を示した。類似度データは、非計量多次元尺度構成法（ｎＭＤＳ）により、低次元数の構造を持つことが明らかにされた。モデルから導かれた類似度と人間の知覚評定に基づく類似度のいずれにおいても２次元が抽出され、それらは、快／不快と覚醒／非覚醒にほぼ対応することがわかった。これらの知見から、顔の内部表現に関する理論的考察と顔表情の認識システムの構築への指針について論じる。

キーワード　顔の表情の知覚、Gabor wavelets、ヒューマンインタフェース、大脳視覚領野

Gabor Wavelet Representation of Facial Expression

Michael J. Lyons[1], Miyuki Kamachi[2], Jiro Gyoba[2], and Shigeru Akamatsu[1]

[1] ATR Human Information Processing Research Laboratories
2-2 Hikaridai Seika-cho Soraku-gun Kyoto 619-02 Japan
[2] Psychology Department, Kyushu University
E-mail : {mlyons, akamatsu}@hip.atr.co.jp
{kamaltm,gyobaltm}@mbox.nc.kyushu-u.ac.jp

Abstract　A facial coding system involving Gabor wavelets was studied was as a potential model of facial expression perception. Gabor wavelet filters were used to model the coding of facial images in visual cortex. Human facial expression perception was studied using a semantic differential paradigm. Measured and modelled facial expression similarity values showed a highly significant degree of correlation. A low-dimensional embedding of the similarity data was discovered using non-metric multidimensional scaling (nMDS). For both model and data, the two recovered dimensions correspond approximately with the stimulus properties pleasant/unpleasant and arousal/non-arousal. Consequences of the findings for theories of facial expression representation and for the design of facial expression recognition systems are discussed.

key words　Facial expression perception; Gabor wavelets; Human-computer Interfaces; Visual Cortex

1. Introduction

Several types of perceptual cue to emotional state are displayed by the face: relative displacements of features (for example, raised eyebrows or an open mouth); local texture changes (for example, furrowing of the brow or creasing of the skin at the corners of the eyes); the signature of the dynamics of these changes; changes in skin hue due to blood flow (for example, blushing). This paper considers the first two of the aforementioned cues explicitly (feature displacement and textural changes). The third type of cue, motion, is considered implicitly through the comparison of relative static configurations. The fourth cue, color, is not considered here.

There is an extensive body of prior work on the psychology of facial expression which we do not attempt to review here. Likewise a growing number of engineering studies [1,2] concern algorithms for automatic facial expression classification, a practical problem with implications for future human-computer interface technology. The project reported on here lies at the interface between psychology and engineering studies. We are investigating a model for human facial expression processing which is motivated by neurobiology of the visual system. In the present paper we report novel findings comparing the behaviour of the model with psychological data and discuss the implications of the model for automatic facial expression recognition systems.

The approach taken here is partly phenomenological. The general framework for representing facial images is based on a geometrical representation of facial structure combined with localized spatial filters which approximate measured attributes of primary visual cortex (V1) receptive fields. This representation is high-dimensional and redundant. A dimensionality reduction technique, non-metric multidimensional scaling (nMDS), is used to search for salient low dimensional structure in the V1-based representation. This method, introduced by Lyons and Morikawa [8] to model facial similarity perception, aims to uncover the nature of the high-level representation of faces.

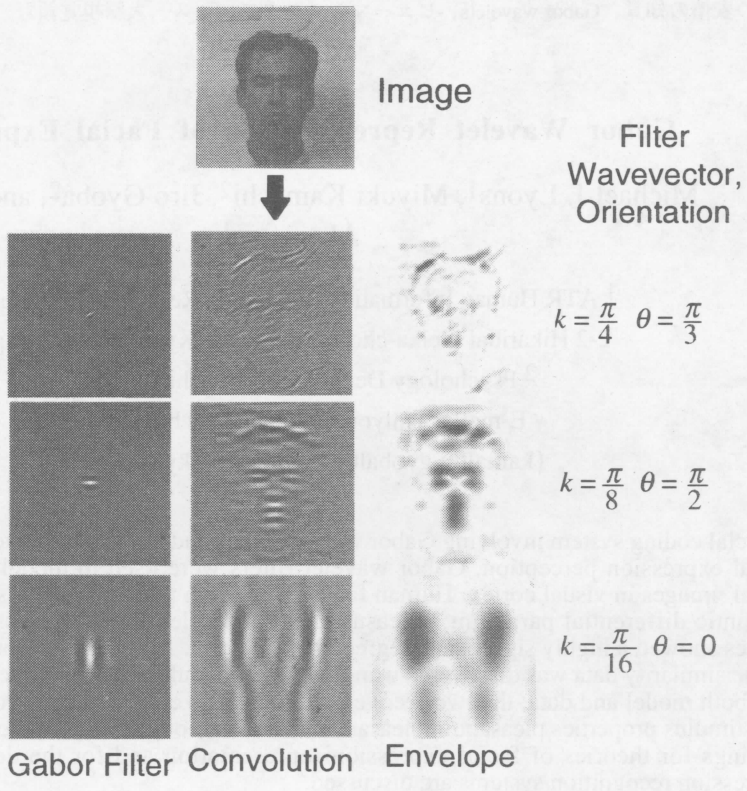

Figure 1 Gabor coding of a facial image.

2. The Model

The Gabor similarity measure used in this work is based on a simple model of primary visual cortex complex cell receptive fields (Fig. 1). The spatial pattern of pixel values centered on image position, $\vec{r}$, is represented by a population response vector, $\vec{R}(\vec{r})$.The components of $\vec{R}(\vec{r})$ are calculated by convolving the image, $I(\vec{r})$, with a set of Gabor wavelet filters, $G_{\vec{k},+}$ and $G_{\vec{k},-}$, which approximate V1 simple cells having even and odd receptive fields respectively and spatial frequency and orientation tuning determined by the wavevector index, $\vec{k}$ [4,6]:

$$R_{\vec{k},\pm}(\vec{r}_0) = \int G_{\vec{k},\pm}(\vec{r}_0,\vec{r})I(\vec{r})d\vec{r}$$

$$G_{\vec{k},+}(\vec{r}) = k^2(e^{-|\vec{r}-\vec{r}_0|^2/2\sigma^2}\cos(\vec{k}\cdot(\vec{r}-\vec{r}_0)) - e^{-\sigma^2/2})$$

$$G_{\vec{k},-}(\vec{r}) = k^2 e^{-|\vec{r}-\vec{r}_0|^2/2\sigma^2}\sin(\vec{k}\cdot(\vec{r}-\vec{r}_0)).$$

The integral of the cosine Gabor filter, $e^{-\sigma^2/2}$ is subtracted from the filter to render it insensitive to the absolute level of illumination. The sine filter does not depend on the absolute illumination level. Three spatial frequencies were used with wavenumbers $(\pi/4, \pi/8, \pi/16)$ measured in inverse pixels. The highest frequency is set at half the Nyquist sampling frequency. The frequency levels are spaced in octaves; $\sigma = \pi$ was used in all calculations, giving a filter bandwidth of about an octave, independent of the frequency level. Six wavevector orientations were used, equally spaced at intervals of 30°, from 0° to 180°. These are the same Gabor wavelet filters described in [7], however filter parameters were chosen to reflect biological values [4,6], while those in [7] optimized performance of a face recognition system.

The components of the Gabor vector, $R_{\vec{k}}$, are defined as the amplitude of the combined even and odd filter responses $R_{\vec{k}} = \sqrt{R_{\vec{k},+} + R_{\vec{k},-}}$ The response amplitude is less sensitive to position changes than are the individual filters and approximates the output of non-endstopped complex cells in V1.

To compute facial similarity, responses of cells having the same spatial frequency and orientation preference are compared at corresponding points in the two facial images. The normalized dot product is used to quantify the similarity of two Gabor response vectors. The V1 similarity of two facial images is the average of the Gabor vector similarity over all corresponding facial points. Since Gabor vectors at neighboring pixels are highly correlated and redundant, it is sufficient to calculate the average on a sparse grid covering the face (Fig. 2). Gabor similarity measures the correlation of Gabor filter responses at like facial position, spatial frequency, and orientation. This similarity measure is related to the one used in [7] but it does not include a term depending explicitly on facial grid geometry.

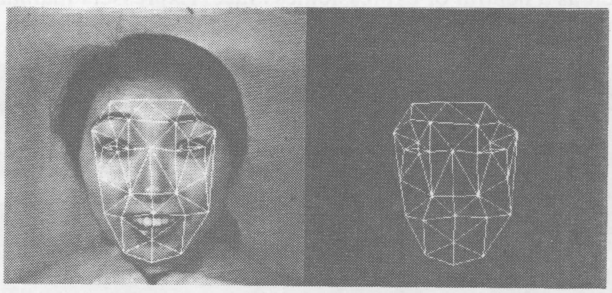

Figure 2 Fiducial grid used to represent the facial geometry.

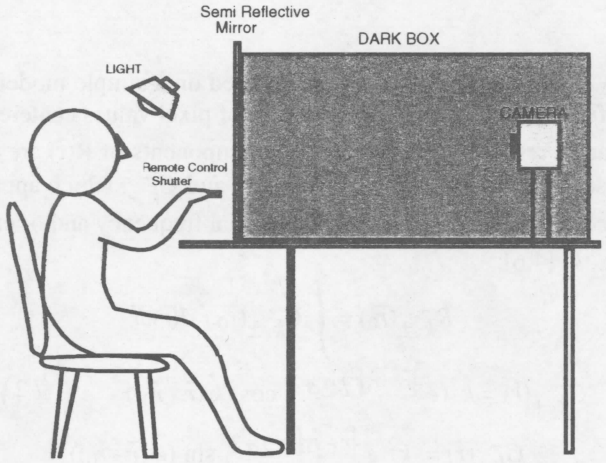

Figure 3 Apparatus used to photograph facial expression stimuli.

The model assumes position and size constancy have been achieved. The face selective cells of IT have low sensitivity to translation and scale changes of the stimulus. Biological mechanisms for achieving correspondence between image and model domains have been discussed elsewhere. In this study, facial graphs were positioned by hand.

A similarity measure based on facial grid geometry was calculated for comparison with the performance of the Gabor wavelet based measure. The distance of each grid point from the node at the tip of the nose was coded as a component in a 34 dimensional vector. Dissimilarity between two grid configurations was taken as the Euclidean distance between vectors. This is a measure of the distortion of the grid due to change in expression.

3. Methods

3.1 Facial Expression Stimuli

Figure 3 shows the apparatus used to collect images of facial expressions. Each expressser photographed herself while looking towards the camera through a semi-reflective sheet of plastic. The face was illuminated evenly with two tungsten lights positioned to each side. The space between the camera and semi-reflective mirror was enclosed with a box to reduce back-reflection of light to the camera and to increase visibility of the reflected face. Expressers were instructed to pose facial expressions so the face would appear as though experiencing a strong emotion. Ten Japanese females (Kyushu University senior undergraduates or graduate students aged 21-26) posed 3 or 4 examples of each of the six fundamental facial expressions as well as a neutral or expressionless face. Expressers maintained a frontal pose of the head towards the camera. Hair was tied back from the face.

Figure 4 Examples of facial expression stimuli used in the experiment. Two examples of subject KA posing each of the six fundamental facial expressions and a neutral, expressionless face are shown.

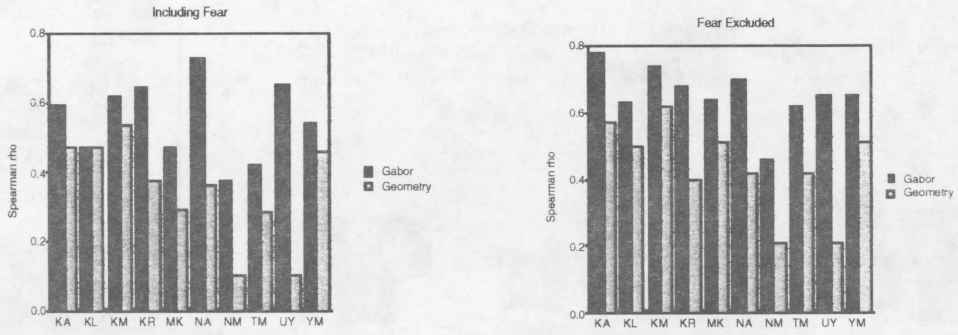

Figure 5. Spearman rank correlation values of similarities computed using model and from human rating data. Left: fear stimuli and ratings excluded. Right: including fear stimuli and ratings.

3.2 Semantic Differential Rating of Images

Subjects rated the images for degree of each of five or six component expressions. Ninety-two Japanese female subjects (Kyushu University undergraduates, aged 18-21) took part in the study. The subject pool was divided into four groups: 1.A, 1.B, 2.A, and 2.B. Group 1.A (15 subjects) rated 94 images (fear expressions excluded) for degree of each of five basic facial expressions (happiness, sadness, surprise, anger, disgust). Group 2.A (15 subjects) rated a different set of 93 images (fear expressions excluded) for degree of the same five basic facial expressions. The combined set of 187 images seen by groups 1.A and 1.B comprise all stimuli excluding images of fear expressions. Group 2.A (31 subjects) rated 108 images for degree of each of the six basic facial expressions (happiness, sadness, surprise, anger, disgust, fear). Group 2.B (31 subjects) rated 111 images for degree of each of the six basic facial expressions. Taken together the 219 images seen by groups 2.A and 2.B comprise the entire stimulus set.

A five point intensity scale was used to quantify each component expression. For example a rating of 1 for the sadness component signifies that the expression appears "not sad at all", while a rating of 5 for the happiness component indicates the expression appears "very happy". Rating values were averaged over all subjects so that images were labelled with two average rating vectors: one with five components, which excludes a fear rating, and one with six components, including averaged ratings on all of the six basic facial expressions. Fear stimuli were rated on six components only. The Euclidean distance between averaged rating vectors was used as a measure of the dissimilarity between images of a given expresser. The negative of the distance was used as a similarity measure for comparison with the model.

4. Results and Discussion

Figure 5 shows bar charts of the Spearman rank correlation (rho) between similarities calculated using the model and from the human ratings. With fear stimuli and ratings excluded, Spearman's rho values ranged from 0.624 (expresser TM) to 0.782 (expresser KA). The average rank correlation between Gabor model and data was 0.679. For the geometry based control similarity measure, the rank correlation between model and data ranged from 0.206 (expresser UY) to 0.619 (expresser KM). The average value of rho was 0.462. Expresser NM was considered to be an outlier and excluded from the above quoted ranges and averages. On inspection of the stimuli, NM's expressions were difficult to interpret in terms of the six basic facial expressions. The two-sided significance of the deviation of rho from 0 indicated high significance for all values of rho calculated.

With fear stimuli and ratings included, Spearman's rho values for correlation between Gabor model and human data ranged from 0.423 (expresser TM) to 0.725 (expresser NA). The average rank correlation between Gabor model and data was 0.568. For the geometry based control similarity measure, the rank correlation between model and data ranged from 0.074 (expresser UY) to 0.527 (expresser KM). The average value of rho was 0.366. Expresser NM was again excluded from the quoted ranges and averages.

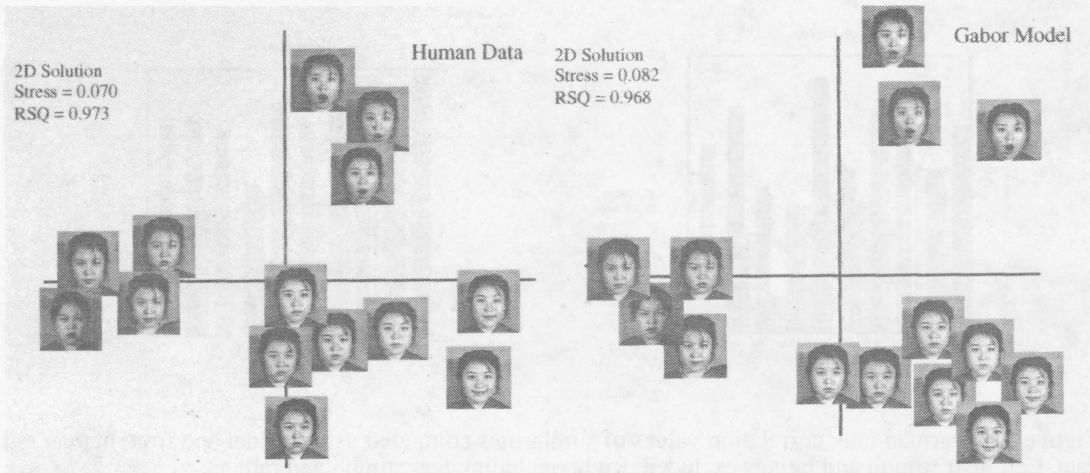

Figure 6 Two dimensional similarity spaces of human and Gabor similarity ratings obtained with non-metric multidimensional scaling analysis. These results are from a pilot study which used the same methods as described in sections 2 and 3.

Similarity matrices calculated from the Gabor-based code and human ratings were analyzed using non-metric multidimensional scaling (nMDS), using the ALSCAL algorithm. nMDS embeds the stimuli in a Euclidean space of specified dimensionality in such a way that the distances between stimuli preserves the rank ordering of the original similarity data. Stress is a measure of the residual badness of fit of the Euclidean distances to the similarity data, while the rsq parameter is the squared linear correlation between stimulus distances in the solution space and dissimilarity values. By monitoring the values of the stress and rsq as the dimensionality of the embedding space was increased it was found that two dimensions were necessary and sufficient to provide a satisfactory fit to the similarity data, for both model and human rating data. Figure 6 shows sample nMDS solutions for human rating similarities and Gabor similarities. Examination of the figure suggests that one (approximately horizontal, in these plots) dimension corresponds to degree of pleasantness versus unpleasantness present in the expression (e.g. happiness on the right side as opposed to anger and disgust on the left side of the axis). An approximately orthogonal (vertical in these plots) axis corresponds to degree of arousal versus non-arousal (surprise above versus sadness below). The same axes were recovered for both model and human rating similarities. In addition, the configuration of individual stimuli along the two axes is quite similar in the two nMDS solutions. Similar results were obtained for some of the other expressers. nMDS solutions obtained when fear stimuli and ratings were included were more variable and showed less agreement between Gabor model and human rating data. A complete set of these results will be included in a future publication.

5. Conclusions

The present study examines a representation of facial expression images based on a multi-scale Gabor wavelet transform aligned with a facial grid. Because Gabor wavelet filters approximately model receptive fields of V1 cells, this may be considered as a model of the front-end processing of facial images by the human visual system. The similarity space generated by the model was compared with human ratings of facial expression images. In all cases the level of correlation between the Gabor wavelet based model and human ratings data was significant and higher than that obtained with a control similarity measure based solely on the geometry of the facial grid. This is not surprising: while grid geometry measures facial distortion due to expression it does not include a description of textural changes. Moreover, a geometry based representation is expected to be quite sensitive to noise in facial grid node positioning, while use of the Gabor wavelet transform amplitude avoids this problem.

Agreement between the Gabor wavelet based model and human ratings was closer when fear stimuli were excluded. While a more in-depth analysis of the data and further experimentation will be necessary to better understand this result, one hypothesis is that fear expressions are processed differently from the other basic facial expressions. Supporting evidence for this hypothesis may be found in the study of amygdala damaged patients by Adolphs et al. [3].

An interesting finding of this study is the agreement between the low-dimensional structure of the Gabor based and human ratings similarity spaces since the former is a function solely of image properties and the latter is derived from semantic labelling data. It is not clear why this should be the case for a simple model that does not consider higher-level facial expression processing. We conjecture that the high-level representation of facial expressions may preserve topological properties of the input (image) space.

Finally, the results indicate that it would be possible to discriminate facial expressions in single images using the output of Gabor wavelet filters aligned with a facial grid. Hence this representation could be used as the front end for an automatic facial expression classification system. As shown in the nMDS analysis, clusters in the similarity space of the Gabor wavelet representation correspond to the different basic facial expressions and may be learned from example images in an unsupervised fashion. That the learned clusters exhibit similar dimensional structure to those found from human semantic rating data may be advantageous for some applications.

Acknowledgments

We thank Phi Tran (University of Southern California) for assistance in the early stages of this project and Reiko Kubota (Kyushu University) for help with the data collection.

References

[1] *International Workshop on Automatic Face and Gesture Recognition, Proceedings*, Zurich, June 26-28, 1995.
[2] *Second International Conference on Automatic Face and Gesture Recognition, Proceedings*, Vermont, October 14-16, 1996.
[3] R. Adolphs, D. Tranel, H. Damasio, & A. Damasio (1994). Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. *Nature*, 372, 669-672.
[4] J. G. Daugman (1985). Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of the Optical Society of America A*, 2, 1160-1169.
[5] P. Ekman and W. V. Friesen (1984). *Unmasking the Face. A guide to recognizing emotions from facial clues* Palo Alto: Consulting Psychologists Press.
[6] J. P. Jones & L. A. Palmer (1987). An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, 58, 1233-1258.
[7] M. Lades, J. C. Vorbruggen, J. Buhmann, J. Lange, C. von der Malsburg, R. P. Wurtz, & W. Konen (1993). Distortion Invariant Object Recognition in the Dynamic Link Architecture, *IEEE Transactions on Computers*, 42, 300-311.
[8] M. J. Lyons & K. Morikawa (1996). A Model Based on V1 Cell Responses Predicts Human Perception of Facial Similarity. *Investigative Opthalmology and Visual Science*, 37, 910.